

An Efficient and Optimal Filter for Identifying Point Sources in Millimeter/Sub-Millimeter Wavelength Sky Maps

T. A. Perera

tperera@iwu.edu

J. R. Schaar

A. Mancera

Department of Physics, Illinois Wesleyan University, Bloomington, IL 61702

G. W. Wilson

Department of Astronomy, University of Massachusetts Amherst, Amherst, MA 01003

K. S. Scott

National Radio Astronomy Observatory, Charlottesville, VA 22903

J. E. Austermann

Center for Astrophysics and Space Astronomy, Boulder, CO 80309

ABSTRACT

A new technique for reliably identifying point sources in millimeter/sub-millimeter wavelength maps is presented. This method accounts for the frequency dependence of noise in the Fourier domain as well as non-uniformities in the coverage of a field. This optimal filter is an improvement over commonly-used matched filters that ignore coverage gradients. Treating noise variations in the Fourier domain as well as map space is traditionally viewed as a computationally intensive problem. We show that the penalty incurred in terms of computing time is quite small due to casting many of the calculations in terms of FFTs and exploiting the absence of sharp features in the noise spectra of observations. Practical aspects of implementing the optimal filter are presented in the context of data from the AzTEC bolometer camera. The advantages of using the new filter over the standard matched filter are also addressed in terms of a typical AzTEC map.

Subject headings: Data Analysis and Techniques, Astrophysical Data

1. Introduction

The discovery and study of sub-millimeter galaxies, or SMGs, has become a key enterprise within millimeter/sub-millimeter (mm/sub-mm) astronomy over the past decade or so (see review by Blain et al. 2002). With bolometric luminosities $\gtrsim 5 \times 10^{12} L_{\odot}$ and star formation rates $\gtrsim 100 M_{\odot} \text{ yr}^{-1}$, SMGs represent some of the most luminous galaxies in the early Universe. The strong negative k -correction at $\lambda > 500 \mu\text{m}$ means that a galaxy of a given luminosity will be equally detectable in a flux-limited survey from $1 \lesssim z \lesssim 10$. A deep, wide-area survey at mm/sub-mm wavelengths is thus a sensitive probe of starburst galaxies from the epoch at which they first turn on through the peak of star formation activity in the Universe at $z \approx 1 - 2$. Several papers have reported on the source counts of SMGs detected at $\lambda = 250 - 2000 \mu\text{m}$ (e.g. Coppin et al. 2006; Bertoldi et al. 2007; Weiß et al. 2009; Vieira et al. 2010; Scott et al. 2012, and references therein), which provide strong constraints on the evolutionary history of massive galaxies (e.g. Baugh et al. 2005; Valiante et al. 2009; Béthermin et al. 2012). SMG surveys from single-dish telescopes also provide a catalog of interesting targets to follow-up with higher angular resolution imaging of the dust and molecular gas in these objects, which inform on the physical processes that trigger and maintain the starbursts in these galaxies.

An important criterion for the success of such surveys is the design of a data analysis scheme that can reliably identify SMG candidates on sky maps. It is especially important to minimize the rate of false detections because they can result in a waste of valuable time and resources in follow-up observations. But reliably identifying astronomical signal in mm/sub-mm maps is a difficult task because these maps are inherently low in signal/noise due to atmospheric contamination, instrument noise, and the high confusion limit of surveys from typical mm/sub-mm telescopes (~ 10 m diameter). In the wide field surveys designed to detect SMGs, they usually appear as point sources much smaller than the angular resolution of the telescopes used. A simple approach to identify point sources that is used by standard data analysis packages (see Stetson 1987, for example) is: 1) smooth the signal map by convolving it with the point spread function (PSF); 2) assume that the errors associated with pixels are uncorrelated and propagate them through the convolution; and 3) select the peaks in the smoothed map based on the per-pixel signal/noise determined from steps 1) and 2). This reduces high-frequency signal variations between adjacent pixels and thus increases the signal/noise for point-source detection.

However, there are two general problems with real mm/sub-mm maps that renders the above procedure inadequate for reliably identifying point sources. First, the noise in these maps is generally not white; it is more pronounced on larger scales or at low (spatial) frequencies. This effect, which gives rise to pixel-pixel correlations, is due mainly to $1/f$ drifts in atmospheric and/or instrumental conditions. In mm/sub-mm astronomy, this problem is usually treated with a “matched filter” implemented in the Fourier domain (e.g. Tegmark & de Oliveira-Costa 1998; Barreiro et al. 2003; Vio et al. 2004; Barnard et al. 2004; Perera et al. 2008; Chapin et al. 2011). However, this filtering technique is optimal only in the case of *uniform* coverage, which brings up the second problem with real mm/sub-mm maps: the coverage of a field is non-uniform in general. Because these

observations are usually carried out by scanning an array of detectors across the field of interest, variations in atmospheric conditions or detector noise during the scan are often responsible for this non-uniformity. With commonly used schemes such as raster- or Lissajous-scanning, coverage also tends to decrease smoothly from the center to the edges of the map.

The correct way to deal with both these problems involves the construction and inversion of a pixel-pixel noise covariance matrix. This path can be extremely challenging computationally for maps containing $\sim 100,000$ pixels. Therefore, a common strategy is to pick out a region of the map that has essentially uniform coverage and then apply the standard matched filter to it. Even when a near-uniform region exists, ignoring small coverage variations within it can have a noticeable effect, as we will show below. Furthermore, despite gradients, the coverage may be deep enough near the edges of the map to identify bright sources with high significance, and these “border” regions can often cover a significant area compared to the near-uniform region. In some cases, significant coverage gradients are unavoidable due to difficult observing conditions or an observing strategy where a mosaic of small maps are stitched together (for example, see Borys et al. 2003; Coppin et al. 2006; Austermann et al. 2010). A truly optimal analysis would help improve the reliability with which point-source candidates are identified in *all* regions of a map despite coverage gradients.

In this paper, we present an algorithm that addresses non-uniform coverage as well as low-frequency noise in a computationally efficient manner. This technique is designed to be optimal in the regime where blending of resolved sources is negligible. The work presented here is an extension of the standard data reduction tools used for point-source extraction in maps taken with the 1.1 mm bolometer array camera, AzTEC (Wilson et al. 2008). While this builds on the existing AzTEC reduction pipeline (Scott et al. 2008; Perera et al. 2008), the principles are generic to any scan-data from mm/sub-mm arrays. In section 2 we present the formal methodology that will be used to identify point sources and to estimate their brightness. Essentially, this method involves a least-squares fit of the PSF to every point on the map. In section 3, we demonstrate that, for the case of uniform coverage, the fit paradigm converges to the familiar and computationally efficient matched filter routinely used in point-source searches. In section 4, we develop an implementation of the pixel-by-pixel fit that can efficiently handle the case of non-uniform coverage. Then, in section 5, we discuss the practicalities of applying this technique to AzTEC data. Finally, in section 6, we conclude by discussing the performance and scope of the optimal filter introduced here.

2. The Method

We leverage our knowledge of the shape of isolated point sources in our maps to construct an *optimal* or *wiener* filter using the point spread function (PSF) as the template. Applying this filter to the map is formally equivalent to centering the PSF on each pixel of the map and fitting for the best-fit amplitude (Stetson 1987). In our case, we store the best fit amplitude at each pixel in a separate map which we will refer to as the *filtered signal map*. In addition, we construct a new

map of the same dimensions as the signal map which contains $1/\text{error}^2$ estimates of the best fit amplitudes. This map will be referred to as the *filtered coverage map*. These maps, augmented by noise realizations of the field, are the primary inputs to the identification of point sources.

2.1. AzTEC data as an example

AzTEC is a 144-element semiconductor-type bolometer array that imaged the sky at 1.1-mm wavelength over two successful observing campaigns: one using the 15-m James Clark Maxwell Telescope (JCMT) in Hawaii from 2005-2006 (Wilson et al. 2008), and the other using the 10-m Atacama Sub-millimeter Telescope Experiment (ASTE) in the Atacama desert of Chile from 2007-2008 (Ezawa et al. 2008). Over these two observing runs, close to a hundred fields were mapped, and each map comprises 100,000 pixels or more. Therefore, computational speed was an important consideration when developing this algorithm. Computational speed will be even more important for newer and upcoming mm/sub-mm observatories like the Large Millimeter Telescope and CCAT that can image the sky at even faster rates with higher angular resolution.

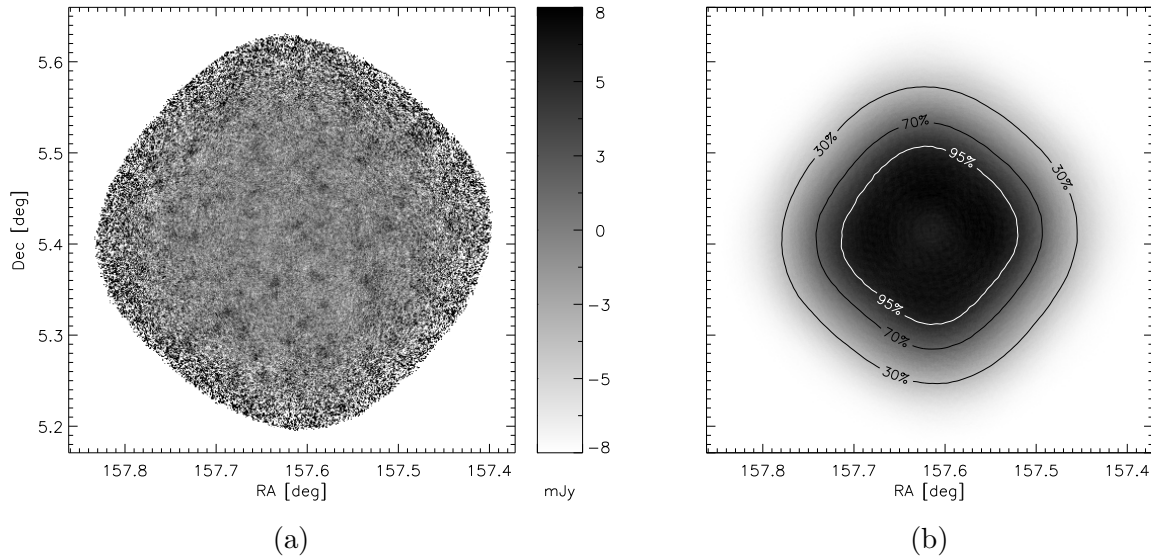


Fig. 1.— (a) 1.1mm-wavelength map of a region surrounding the quasar SDSSJ1030+0524, obtained with AzTEC-ASTE. The outer regions are saturated high (black) or low (white) on an astronomically interesting scale (± 8 mJy). Astronomical features are visible in the central unsaturated (grey scale) region, mainly as PSF-sized dark patches (“blobs”), or point sources. (b) Coverage map of this field. It is filled with the values $1/\epsilon^2(\mathbf{x}_p)$, where $\epsilon(\mathbf{x}_p)$ is the uncertainty of the signal at the p ’th pixel of (a). Coverage contours are labeled relative to the maximally covered (dark) central regions.

To illustrate the starting point for the filtering process, we present in Fig. 1 an unfiltered AzTEC signal map and the corresponding coverage map. In the AzTEC data analysis, coverage maps have units of $1/\text{Jy}^2$ and are filled with estimates of $1/\text{error}^2$ at each pixel. In Fig. 1 and throughout the rest of this paper, we will use maps and plots related the AzTEC-ASTE data on a region centered around the high-redshift quasar SDSSJ1030+0524 to illustrate our points. This data set consists of 45 observations of the field, each lasting about forty minutes, carried out over November 2008. Scientific results from this AzTEC-ASTE field are published in Humphrey et al. (2011) and Zeballos (in prep.). These data were taken by scanning the AzTEC array in a Lissajous pattern, and results in a map with deepest coverage in the center, and decreasing coverage toward the edges of the map.

2.2. Assumptions

Two assumptions, both of which are generally applicable to most observations, are critical for the efficient implementation of the process described below. They are:

1. *The effective PSF does not change with location on the map.* We represent the PSF by the two-dimensional function $f(\mathbf{x})$. The value of this function at the i 'th pixel when the PSF is centered on the p 'th pixel will be denoted by $f(\mathbf{x}_i - \mathbf{x}_p)$, where \mathbf{x}_i and \mathbf{x}_p are 2-d vectors that specify the locations of the i 'th and p 'th pixels, respectively. We note that while this assumption is explicitly not valid for some instruments (e.g., the *Chandra* X-ray telescope), this is a reasonable assumption for many imaging instruments.
2. *Noise-induced correlations between two pixels, hereafter denoted $\text{Corr}(\mathbf{x}_k, \mathbf{x}_l)$, depend only on the distance $|\mathbf{x}_k - \mathbf{x}_l|$, between the pixels.* That is, neither the location of the two pixels on the map nor their relative orientation determines the correlation in their noise. In such cases, $\text{Corr}(\mathbf{x}_k, \mathbf{x}_l)$ may be conveniently expressed as the inverse Fourier transform (IFT) of the *power spectral density* (PSD) of noise in the map. If we denote the PSD by $V^2(\mathbf{k}_a)$, where \mathbf{k}_a is a 2-d vector in the Fourier domain of the map,

$$\text{Corr}(\mathbf{x}_k, \mathbf{x}_l) = \sum_a^{N_{\text{pixel}}} V^2(\mathbf{k}_a) \exp[2\pi j \mathbf{k}_a \cdot (\mathbf{x}_k - \mathbf{x}_l)], \quad (1)$$

where N_{pixel} is the total number of pixels in the map, and the sum is over all N_{pixel} vectors \mathbf{k}_a . (Appendix A contains a brief description of the discrete Fourier transform conventions and their corollaries used here.)

In Eq. 1, $V^2(\mathbf{k}_a)$ is a normalized form of the PSD that satisfies $\sum_a^{N_{\text{pixel}}} V^2(\mathbf{k}_a) = 1$, so that the diagonal elements of the correlation matrix evaluate to 1. Note that for a completely flat PSD ($V^2(\mathbf{k}_a) = \text{cnst}$) the correlation matrix is diagonal. In most observations the noise PSD varies smoothly with \mathbf{k} and has *broad* features rather than narrowly peaked ones (see Fig. 3, for example). In such cases the correlation matrix is band-diagonal (i.e., $\text{Corr}(\mathbf{x}_k, \mathbf{x}_l) \simeq 0$ when $|\mathbf{x}_k - \mathbf{x}_l|$ is large).

This is a property that we will exploit below in our pursuit of computational efficiency.

2.3. An optimal filter via a generalized least-squares fit

We implement the optimal filter by minimizing the quantity

$$\chi_p^2 = \sum_{k,l=0}^{N_{\text{pixel}}} [d(\mathbf{x}_k) - s_p f(\mathbf{x}_k - \mathbf{x}_p)] W_{kl} [d(\mathbf{x}_l) - s_p f(\mathbf{x}_l - \mathbf{x}_p)], \quad (2)$$

for each pixel, p , in the map where $d(\mathbf{x}_k)$ and $d(\mathbf{x}_l)$ denote the value of the unfiltered signal map at the k 'th and l 'th pixels, s_p is the amplitude of the fitted PSF, and W is an $N_{\text{pixel}} \times N_{\text{pixel}}$ *weight* matrix whose calculation is described below. The summations in Eq. 2 are over all pixels of the map. Eq. 2 is minimized when

$$s_p = \frac{\sum_{k,l} W_{kl} d(\mathbf{x}_k) f(\mathbf{x}_l - \mathbf{x}_p)}{\sum_{k,l} W_{kl} f(\mathbf{x}_k - \mathbf{x}_p) f(\mathbf{x}_l - \mathbf{x}_p)}. \quad (3)$$

For least-squares fits, the minimum variance choice of the weight matrix is $W = C^{-1}$ where C is the pixel-pixel noise covariance matrix (i.e. the map's covariance in the absence of signal). In this case, the χ_p^2 calculated in Eq. 2 is drawn from a true χ^2 distribution with $N_{\text{pixels}} - 1$ degrees of freedom. The elements of C are

$$C_{kl} = \epsilon(\mathbf{x}_k) \text{Corr}(\mathbf{x}_k, \mathbf{x}_l) \epsilon(\mathbf{x}_l) = \epsilon(\mathbf{x}_k) \left(\sum_a V^2(\mathbf{k}_a) e^{2\pi j \mathbf{k}_a \cdot (\mathbf{x}_k - \mathbf{x}_l)} \right) \epsilon(\mathbf{x}_l) \quad (4)$$

where $\epsilon(\mathbf{x}_k)$ and $\epsilon(\mathbf{x}_l)$ are the standard deviations of noise in the k -th and l -th pixels respectively. Note that the diagonal elements, C_{kk} evaluate to $\epsilon^2(\mathbf{x}_k)$, as they should.

The elements of the weight matrix can then be calculated as

$$W_{kl} = [C^{-1}]_{kl} = \frac{1}{N_{\text{pixel}}^2} \frac{1}{\epsilon(\mathbf{x}_k)} \left(\sum_a \frac{e^{2\pi j \mathbf{k}_a \cdot (\mathbf{x}_k - \mathbf{x}_l)}}{V^2(\mathbf{k}_a)} \right) \frac{1}{\epsilon(\mathbf{x}_l)}. \quad (5)$$

Using Eq. A5 of Appendix A, it is straightforward to verify that $\sum_l W_{kl} C_{lm} = \delta_{km}$. Substituting the above form of W_{kl} into Eq. 3, we may express the numerator and denominator of Eq. 3 as

$$N_p = \frac{1}{N_{\text{pixel}}^2} \sum_{k,l} \frac{f(\mathbf{x}_k - \mathbf{x}_p)}{\epsilon(\mathbf{x}_k)} \left(\sum_a \frac{e^{2\pi j \mathbf{k}_a \cdot (\mathbf{x}_k - \mathbf{x}_l)}}{V^2(\mathbf{k}_a)} \right) \frac{d(\mathbf{x}_l)}{\epsilon(\mathbf{x}_l)} \quad (6)$$

$$D_p = \frac{1}{N_{\text{pixel}}^2} \sum_{k,l} \frac{f(\mathbf{x}_k - \mathbf{x}_p)}{\epsilon(\mathbf{x}_k)} \left(\sum_a \frac{e^{2\pi j \mathbf{k}_a \cdot (\mathbf{x}_k - \mathbf{x}_l)}}{V^2(\mathbf{k}_a)} \right) \frac{f(\mathbf{x}_l - \mathbf{x}_p)}{\epsilon(\mathbf{x}_l)}. \quad (7)$$

In practice, we form separate “maps” corresponding to N and D because the latter is useful by itself (see below). In doing so, the nested sums of Eq. 6 and Eq. 7 will be recognized as Fourier transforms, allowing us to benefit from the efficiency of the FFT algorithm. The element-by-element ratio of these two, $s_p = N_p/D_p$, forms the filtered signal map.

2.4. Error propagation

The error in our estimate of s_p , which we denote n_p , is the rms deviation of the best fit PSF amplitude, s_p , from the actual amplitude, a_p , of a point source that is centered on the p ’th pixel. i.e. $n_p^2 = \langle |s_p - a_p|^2 \rangle$. To estimate n_p^2 , we consider the influence of a point source located at the p ’th map pixel on a neighboring pixel i .

$$d(\mathbf{x}_i) = a_p f(\mathbf{x}_i - \mathbf{x}_p) + \sqrt{C_{ii}}. \quad (8)$$

As mentioned above, we recognize that the possibility of other resolved sources being quite close to pixel i (source blending) is ignored in Eq. 8 and must be treated iteratively. Using Eq. 8 for $d(\mathbf{x}_i)$ and Eq. 3 for s_p , our estimate for n_p^2 is

$$n_p^2 = \langle |s_p - a_p|^2 \rangle = \frac{\sum_{k,l} \sum_{t,u} W_{kl} W_{tu} C_{kt} f(\mathbf{x}_l - \mathbf{x}_p) f(\mathbf{x}_u - \mathbf{x}_p)}{\left(\sum_{k,l} W_{kl} f(\mathbf{x}_k - \mathbf{x}_p) f(\mathbf{x}_l - \mathbf{x}_p) \right)^2}, \quad (9)$$

and since $W = C^{-1}$, the above expression simplifies to

$$n_p = \frac{1}{\sqrt{\sum_{k,l} W_{kl} f(\mathbf{x}_k - \mathbf{x}_p) f(\mathbf{x}_l - \mathbf{x}_p)}} = \frac{1}{\sqrt{D_p}}. \quad (10)$$

Thus, when the optimal W is used, no extra steps are needed to evaluate n_p , as D_p already exists from generating the filtered signal map.

2.5. Source significance and goodness of fit

The filtered signal-to-noise map, filled with values s_p/n_p , can be used to both identify *and locate* sources of high significance. Rewriting Eq. 2 in a more convenient form using Eq. 3 and Eq. 10 gives

$$\chi_p^2 = \sum_{k,l} W_{kl} d(\mathbf{x}_k) d(\mathbf{x}_l) - (s_p/n_p)^2. \quad (11)$$

Because the first term in Eq. 11 is common to all pixels p , the second term is a direct indication of the goodness of fit at the p ’th pixel. Thus, the locations of peaks in the signal-to-noise map, s_p/n_p , provide the best estimates of point-source positions on the map.

3. The case of uniform coverage

It is instructive to first consider the case of uniform coverage. In this case, $\epsilon(\mathbf{x}_k) = \epsilon(\mathbf{x}_l) = \text{cst}$ in Eqs. 6 and 7 and so can be moved outside the summations. The sums over k and l are then recognized as Fourier transforms (FTs) of the signal map ($d(\mathbf{x})$) and the PSF ($f(\mathbf{x})$), leading to the result that

$$s_p = \frac{\sum_a \frac{\tilde{f}^*(\mathbf{k}_a) \tilde{d}(\mathbf{k}_a)}{V^2(\mathbf{k}_a)} \exp[2\pi j \mathbf{k}_a \cdot \mathbf{x}_p]}{\sum_a \frac{|\tilde{f}(\mathbf{k}_a)|^2}{V^2(\mathbf{k}_a)}}, \quad (12)$$

where $\tilde{f}(\mathbf{k}_a)$ and $\tilde{d}(\mathbf{k}_a)$ are the 2-d FTs of $f(\mathbf{x}_i)$ and $d(\mathbf{x}_i)$ respectively. The above expression is the familiar “matched filter” that is commonly used in mm/sub-mm astronomy for identifying point sources (for example, see Tegmark & de Oliveira-Costa 1998; Barreiro et al. 2003; Vio et al. 2004; Barnard et al. 2004; Chapin et al. 2011) and is essentially a band-pass filter in the sense that the data (\tilde{d}) are low-pass filtered by the PSF \tilde{f} and high-pass filtered by $1/V^2$ as the PSD usually displays a $1/f$ -type trend with spatial frequency. The denominator of Eq. 12 simply provides the correct normalization. What Eq. 12 shows is that the point-source finding technique of fitting a PSF to each map pixel converges to the conventional matched filter in the case of uniform coverage.

Another important feature of Eq. 12 is that, in terms of the FTs $\tilde{f}(\mathbf{k}_a)$ and $\tilde{d}(\mathbf{k}_a)$ the numerator and the denominator are each reduced to a single sum of N_{pixel} terms and the total calculation scales as $N_{\text{pixel}} \log_2 N_{\text{pixel}}$ rather than as N_{pixel}^2 . As for having to evaluate s_p at each pixel, we note that (1) the denominator needs to be evaluated just once for all pixels and (2) a “map” of the numerator is simply an inverse Fourier transform (IFT). Therefore, the entire filtering process can be accomplished with $\sim \mathcal{O}(N_{\text{pixel}} \log_2 N_{\text{pixel}})$ calculations. On a desktop computer with a 2.5 GHz Intel processor and 4.7 GB of RAM, the $\sim 200,000$ pixel map of Fig. 1 can be filtered according to Eq. 12, using the IDL programming language, in under 0.5 seconds.

Using the standard matched filter is similar to performing a simple least squares minimization under the approximation that all pixels in the map have the same uncertainty. If the underlying noise is *not* uniform and this method is used, one must recognize that $W \neq C^{-1}$ and so the error distributions of the s_p cannot be formally calculated as given in Eq. 10 above.

4. The case of non-uniform coverage

We show here that, in a map with non-uniform coverage, it is possible to treat the general problem of point source identification in a computationally efficient, yet mathematically sound way. We find that loosening the assumption of uniform coverage results in an increase in computation time that is noticeable but inconsequential in practical terms. As the calculation of D turns out to be more complex than the calculation of N , we tackle N first.

4.1. The calculation of N

N , as expressed in Eq. 6, can be calculated through a sequence of FTs and IFTs. We motivate this by casting Eq. 6 in a more suggestive manner:

$$N_p = \frac{1}{N_{\text{pixel}}} \sum_k \frac{f(\mathbf{x}_k - \mathbf{x}_p)}{\epsilon(\mathbf{x}_k)} \left(\sum_a \frac{e^{2\pi j \mathbf{k}_a \cdot \mathbf{x}_k}}{V^2(\mathbf{k}_a)} \left(\frac{1}{N_{\text{pixel}}} \sum_l \frac{d(\mathbf{x}_l)}{\epsilon(\mathbf{x}_l)} e^{-2\pi j \mathbf{k}_a \cdot \mathbf{x}_l} \right) \right) \quad (13)$$

Proceeding from the innermost parentheses outward, the steps needed to calculate N are:

1. Form a new map M , filled with values $M(\mathbf{x}_l) = d(\mathbf{x}_l)/\epsilon(\mathbf{x}_l)$.
2. The calculation within the innermost parentheses can be recognized as a Fourier Transform (see Appendix A). Therefore, form $\tilde{M} = \text{FT}[M]$.
3. Similarly, the calculation within the next set of parentheses is an IFT. Thus, form $P = \text{IFT}[\tilde{M}/V^2]$.
4. Form a new matrix Q , which is filled with values $Q(\mathbf{x}_k) = P(\mathbf{x}_k)/\epsilon(\mathbf{x}_k)$. The sum over k in Eq. 13 is simply a convolution of Q by the PSF f .
5. As it is computationally advantageous to carry out this convolution in the Fourier domain, calculate $\tilde{f} = \text{FT}[f]$ and $\tilde{Q} = \text{FT}[Q]$.
6. Finally, $N = \text{IFT}[\tilde{f}^* \tilde{Q}]$.

There are 5 FFTs in this sequence. On the same standard desktop PC described in Section 3, the complete computation sketched above takes less than one second for a $\sim 200,000$ -pixel map.

4.2. The calculation of D

The calculation of the denominator, D , is more involved than the numerator as there are no simplifications that allow D to reduce to a series of Fourier Transforms. As a result, a full calculation of D requires of $\mathcal{O}(N_{\text{pixel}}^2)$ calculations. In the case of the $\sim 200,000$ -pixel map used here, this takes about seven hours on the basic computing platform considered in Section 3. We can do better though. Below we give an approximation to D with an accuracy better than 0.1% that takes ~ 2 minutes to calculate.

Our Approximation of D makes use of the band-diagonal nature of the noise correlation matrix or, equivalently, the smoothly varying nature of $V^2(\mathbf{k}_a)$ as discussed in section 2. We start by defining a new vector, $\mathbf{x}_d = \mathbf{x}_k - \mathbf{x}_l$, and re-writing Eq. 7 as

$$D_p = \frac{1}{N_{\text{pixel}}^2} \sum_d Z(\mathbf{x}_d) \sum_l \frac{f(\mathbf{x}_l - \mathbf{x}_p + \mathbf{x}_d) f(\mathbf{x}_l - \mathbf{x}_p)}{\epsilon(\mathbf{x}_l - \mathbf{x}_d) \epsilon(\mathbf{x}_l)}, \quad (14)$$

where $Z(\mathbf{x}_d)$ is the IFT of $1/V^2(\mathbf{k}_a)$, or

$$Z(\mathbf{x}_d) = \sum_a \frac{e^{2\pi j \mathbf{k}_a \cdot \mathbf{x}_d}}{V^2(\mathbf{k}_a)}. \quad (15)$$

To clearly demonstrate that the sum over l is a convolution, we define the new functions F_d and R_d as follows:

$$F_d(\mathbf{x}) = f(\mathbf{x} + \mathbf{x}_d)f(\mathbf{x}) \quad (16)$$

$$R_d(\mathbf{x}) = \frac{1}{\epsilon(\mathbf{x} + \mathbf{x}_d)\epsilon(\mathbf{x})}. \quad (17)$$

Thus, F_d (or R_d) is the product of f (or $1/\epsilon$) and a version of f (or $1/\epsilon$) that is shifted by the vector \mathbf{x}_d . Using these definitions, we may express D_p as

$$D_p = \frac{1}{N_{\text{pixel}}^2} \sum_d Z(\mathbf{x}_d) \sum_l F_d(\mathbf{x}_l - \mathbf{x}_p) R_d(\mathbf{x}_l), \quad (18)$$

which is recognizable as a convolution. Finally, using the discrete convolution theorem (see Eq. A5), we find that

$$D_p = \frac{1}{N_{\text{pixel}}} \sum_d Z(\mathbf{x}_d) \left(\sum_a \tilde{F}_d^*(\mathbf{k}_a) \tilde{R}_d(\mathbf{k}_a) e^{-2\pi j \mathbf{k}_a \cdot \mathbf{x}_p} \right). \quad (19)$$

The advantage of casting D in the form of Eq. 19 is the following: $Z(\mathbf{x}_d)$ is narrowly peaked near $\mathbf{x}_d = \mathbf{0}$ because it is the IFT of a smoothly varying function $1/V^2(\mathbf{k}_a)$. Therefore, the sum over d converges very rapidly at small \mathbf{x}_d and changes very little at large \mathbf{x}_d and so high accuracy in the calculation of D can still be obtained after truncating the calculation at $N_t \ll N_{\text{pixel}}$ terms. The choice of N_t will depend on the properties of $Z(\mathbf{x}_d)$ and the desired accuracy in D . For example, for our AzTEC maps we find that if we limit the sum over d to the $N_t = N_{\text{pixel}}/200 \approx 1,000$ terms where $|Z(\mathbf{x}_d)|$ is largest, D converges to within 0.1% of its final form in useful parts of the map (the $> 5\%$ coverage region) and requires only 2 minutes to compute.

The steps we follow to generate our approximate D are then

1. Compute $Z = \text{IFT}[1/V^2]$ and find the N_t positions \mathbf{x}_d where $|Z|$ is largest. Then, for each of those \mathbf{x}_d ,
2. shift f by \mathbf{x}_d and multiply by the unshifted f to form F_d . Then generate R_d by performing the same steps on $1/\epsilon$;
3. compute the Fourier Transforms $\tilde{F}_d = \text{FT}[F_d]$ and $\tilde{R}_d = \text{FT}[R_d]$;
4. compute the term within parentheses in Eq. 19 as $G(\mathbf{x}_d) = \text{IFT}(\tilde{F}_d^* \tilde{R}_d)$;
5. Repeat for the N_t vectors (\mathbf{x}_d) chosen, and sum the terms $Z(\mathbf{x}_d)G(\mathbf{x}_d)/N_{\text{pixel}}$.

5. Application of the optimal filter to AzTEC maps

Here, we will demonstrate how the methods developed above have been applied to AzTEC maps, using the field of Fig. 1 as an example. Although the examples and justifications presented here are based on AzTEC data, we note that many of the trends and techniques identified here are typical of most observations and analysis chains.

5.1. Validation of assumptions using the PSF

In AzTEC, we have always used an accurate simulation of the PSF, rather than a generic form such as a Gaussian, as the template $f(\mathbf{x})$ used in the filter. According to the methodology developed in section 2, this is the correct template to use, and should lead to higher accuracy of filtered maps. The simulation used for generating the PSF includes effects such as individual detector beam shapes, the location and orientation of the field during each observation, as well as “cleaning” and filtering steps identical to those used on the true field. The details of how the PSF is generated in the AzTEC data analysis can be found in Downes et al. (2012) and Scott et al. (2008). Fig. 2(a) shows the PSF that applies to the AzTEC map of Fig. 1. Strictly speaking, the

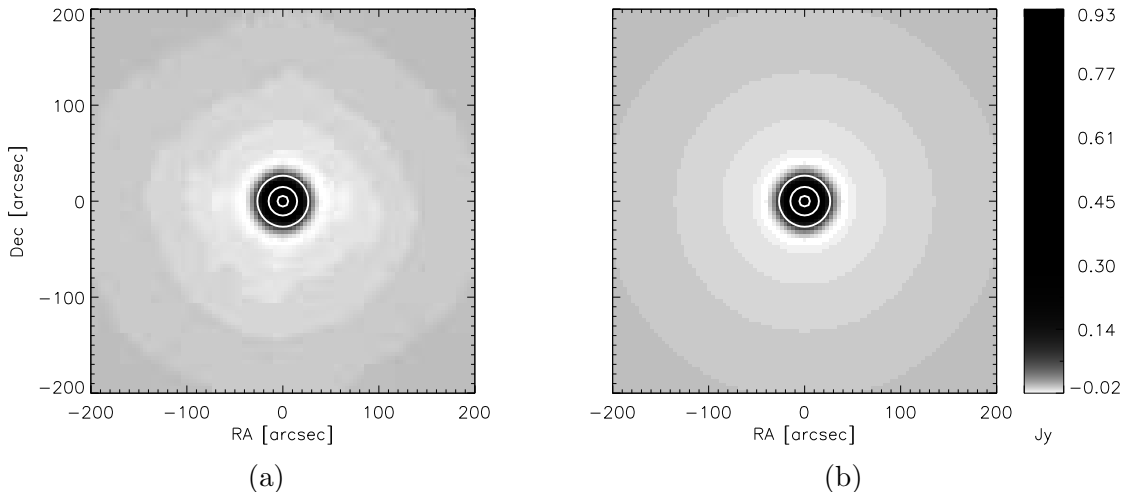


Fig. 2.— (a) The PSF resulting from optics and data “cleaning.” The grey scale is stretched to emphasize the asymmetric and, at times negative, low-signal values. The dark high-signal regions follow a rotationally symmetric profile. The three white contours are at 10%, 50%, and 90% of the peak value. The peak value is 0.93% of the inserted source brightness, indicating the amount by which a point source is attenuated due to the “cleaning” of atmospheric contamination. (b) The rotationally symmetrized version of (a) used as the fit template. The central region has a FWHM of $\sim 29''$.

PSF of Fig. 2(a) only applies to a point source at a particular location on the map. Therefore,

before adopting assumption 1 of section 2.2, we constructed several PSFs that apply to different locations on a map and verified that they are all very similar (at the few percent level) in terms of amplitude, width, and general shape (Scott et al. 2008). The only noticeable difference between these PSFs is the orientation of the low-level asymmetric features seen in Fig 2(a). Therefore, we use a rotationally symmetrized version of the PSF, shown in Fig. 2(b), as the filter template. To evaluate the inaccuracy introduced by rotationally symmetrizing the PSF, we have fit the PSF of Fig. 2(b) to the PSF of Fig. 2(a), and find that the best fit amplitude differs from the actual amplitude of Fig. 2(a) by $\sim 0.5\%$, which is small compared to other errors we expect. Therefore, we adopt assumption 1 and the rotationally symmetrized PSF of Fig. 2(b).

Assumption 2 of section 2.2 would be difficult to adopt if the field is imaged only once, so that the field has an essentially fixed Az-El orientation during the observation and each point on the map is scanned over in essentially one direction. However, AzTEC and other mm/sub-mm surveys routinely adopt assumption 2 because each of the mapped fields is imaged many times with varying orientations and scan directions. This is evident in Fig. 2(a) where the grey scale has been stretched to highlight the low-level asymmetric features of the PSF. We believe that such features are due to a small anisotropy in the final aggregate of scan directions from the 45 observation of this field. However, the fact that these features are so faint compared to the symmetric central parts (refer to the color bar of Fig. 2) supports our adoption of assumption 2.

5.2. Estimation of the PSD

In the AzTEC data analysis, the PSD corresponding to an observation is evaluated using Fourier transforms (FTs) of noise realization maps. Our noise realizations are generated using a “jackknife” method, where the signs of the detector time-streams are switched many times for each observation prior to map making (Scott et al. 2008). An important consideration is to use long-enough time lags between the random sign switchings in order to preserve long-range noise features (low-frequency map noise) due to residual atmospheric contamination and instrumental drifts. The ~ 10 s time scales used are longer than the scan turn-around times but short enough that astronomical signal will not show up in noise realizations. Next, the noise realization time-streams are put through the same cleaning and filtering steps as the actual data, before maps are made. For a map such as Fig. 1, we generate 100 independent noise realization maps.

Because noise realizations are used for generating the PSD, it is free from astronomical signal and only includes the two dominant noise sources, residual atmospheric contamination and instrumental noise. The sub-dominant yet noticeable contribution of confusion noise is left out from the PSD for convenience.¹ However, the method presented here, of using a PSD to char-

¹We note that other analyses that make use of the filtered signal map, such as estimating de-boosted source fluxes, stacking analyses, and number-counts estimates will not be skewed by the omission of confusion noise due to the use of source realizations and/or the AzTEC PSF having zero mean (see Scott et al. 2008). On the other hand, source

acterize noise correlations, can fully accommodate astronomical “noise sources” such as confusion noise. For instance, Chapin et al. (2011) describe how confusion noise was included in their matched filter. In general, astronomical effects such as confusion noise are frequency dependent. At longer wavelengths, a method for including the CMB within a matched filter is described in Tegmark & de Oliveira-Costa (1998).

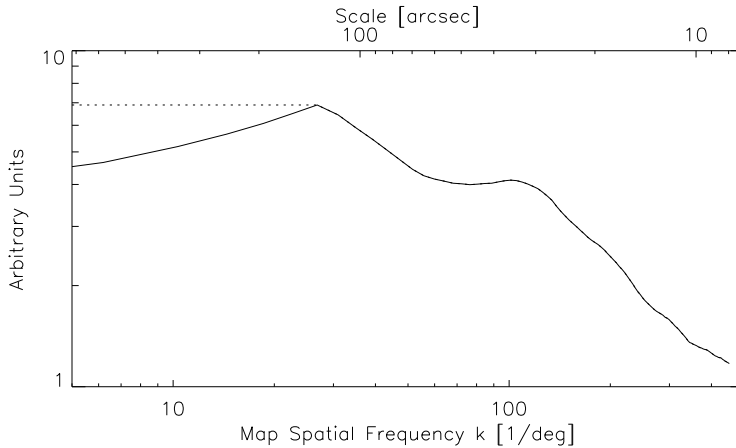


Fig. 3.— The average power spectral density (PSD) of noise in noise-realization maps. The broad peak at the PSF scale ($\sim 30''$) is believed to be caused by atmospheric emission that is imaged in between the random sign switches used to generate noise realizations. The dotted line indicates an alteration to the PSD (of plateauing at low frequency) that we have tried.

In our case, the noise PSD is generated by truncating each noise map to include only the central $> 70\%$ coverage region (see Fig. 1(b)). Then, the 2-d FTs obtained from these regions are rotationally symmetrized and averaged to obtain the PSD shown in Fig. 3. As expected, the noise increases with decreasing k over most of the k range. A broad peak is visible at $\sim 30''$, which is close to the FWHM of the PSF. This, we believe, is caused by the optical imaging of atmospheric fluctuations in between the random sign switches used to generate noise realizations. Near the $\sim 100''$ scale, the PSD turns over and starts to decrease. This decline cannot be a map-size effect because the $> 70\%$ coverage region used to construct the PSD extends $> 700''$ in all directions. We believe that the PSD’s turn-over is real and that it is caused by the principle component analysis (PCA) based cleaning that the data is put through prior to map making, in order to mitigate the effects of atmospheric contamination and detector cross-talk. As PCA cleaning makes use of

lists for follow-up observations, which are based purely on s/n will be very slightly biased due to the omission of confusion noise. This bias would favor peaks found in regions that have an over-density of unresolved sources, which can be beneficial to our understanding of SMGs and their environments.

detector-detector time-stream correlations to “subtract” out these effects, it makes sense that the PSD shows a decline on the scale of the detector array (the spacing of individual detector beams is $\sim 40''$ and the footprint of the entire array is $\sim 480''$ for AzTEC-ASTE). To test if this long-range decline in the PSD is indeed real, we have performed the optimal filter presented here using the measured PSD as well as a modified PSD that plateaus at small k , as indicated by the dotted line of Fig. 3. The fact that there is no perceptible difference in results indicates that the map noise power is indeed low at low k . As noted earlier, an important property of the PSD of Fig. 3 is that it is a smoothly varying function, as opposed to a narrowly peaked one.

5.3. Filtering of maps

Once the PSF and PSD are available, the methods of section 4 can be applied to obtain the filtered signal and coverage maps. In practice, more time is spent on the calculation of N than the calculation of D during the filtering process, even though D involves a more complex calculation (see section 4). This is because N needs to be evaluated separately for the actual map and each noise realization, as N depends on $d(\mathbf{x})$ (see Eq. 6), while D needs to be evaluated only once for all 101 maps. For the example field used, the filtered signal and coverage maps are shown in Fig. 4.

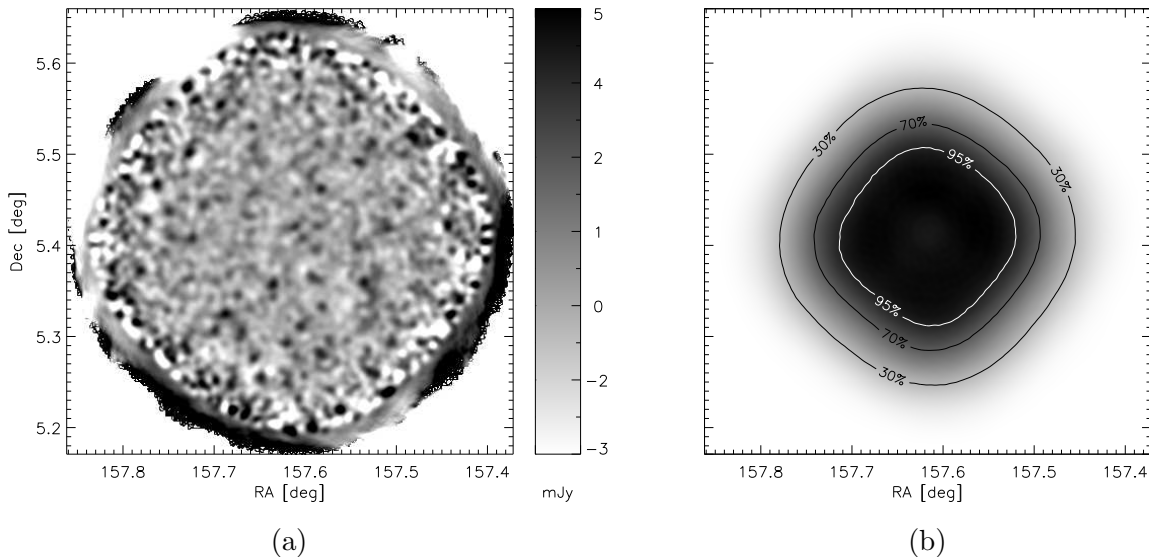


Fig. 4.— Filtered signal map and filtered coverage map of the entire field, generated according to the optimal filter of section 4.

The filtered coverage map must be filled with values $1/n_p^2$, and therefore, according to Eq. 10, it is simply equal to the map D . But D is generated by propagating the initial noise variances, $\epsilon(\mathbf{x}_p)$, according to Eq. 7. Thus the noise estimates contained in D are only as good as the starting noise

estimates $\epsilon(\mathbf{x}_p)$, which are computed from estimates of individual detectors’ noise levels “near the time” that they contribute a sample to the map pixel in question (see Scott et al. 2008, for details). Therefore, before proceeding to the step of identifying point sources, we check the accuracy of noise estimates contained in D using noise realization maps. First, we form a second noise estimate n'_p for each pixel by taking the standard deviation of the 100 noise-map values found at that pixel. Next, we form a second coverage map T filled with the values $1/n_p'^2$. Although T provides a robust estimate of coverage, it is noisier than the original coverage map D due to the sample size of 100 (we expect a $\sim 7\%$ error in T_p for this sample size). Despite the noisy appearance of T , its overall shape agrees very well with D . However, T and D often differ from each other by an overall scaling factor. We determine this scaling factor by comparing the average values of T and D in the $>70\%$ coverage region and applying a correction factor to D in order to obtain the final filtered coverage map. Thus, the filtered coverage map is ηD , where

$$\eta = \frac{\langle T \rangle_{70\%}}{\langle D \rangle_{70\%}}, \quad (20)$$

and η usually lies in the range 0.85-0.95, depending on the field. Although we use just the $>70\%$ coverage region to find η , the agreement between T and ηD remains good out to the very edge of the mapped region until the coverage dips below 5%. Beyond this region, ηD is consistently larger than T . Therefore, we do not trust the filtered coverage map beyond the 5% region and exclude this area when searching for point sources. Finally, in Fig. 5, we present the filtered signal-to-noise map, which is used directly to find point sources. It is generated by dividing the filtered signal map (N_p/D_p) by the filtered noise map ($1/\sqrt{\eta D_p}$).²

6. Results and discussion

It is encouraging that the signal-to-noise map of Fig. 5 has its most prominent peaks near the center, where the coverage is highest, and that the number of peaks as well as their amplitude declines smoothly toward the outer regions. Thus, even though the filtered signal map has large fluctuations near the edges (see Fig. 4(a)), the noise in those regions are accounted for in this optimal filter and there is no need for a by-hand coverage cut to eliminate spurious behavior near the edges. Throughout the AzTEC data analysis campaign, which has led to the publication of many source lists, variations of the filter presented here have been used. Therefore, the effectiveness of this technique may be assessed by the success rate of follow-up observations. In this regard, AzTEC has enjoyed an excellent record thus far. For instance, of 15 AzTEC point sources followed up by the Submillimeter Array (SMA), all were successfully observed (Younger et al. 2007, 2009). This is a very good record in comparison to the usual success rate of mm/sub-mm follow up studies.

²The correction factor η does not need to be applied in generating the filtered signal map because N and D both contain two factors of ϵ (see Eq. 6 and Eq. 7).

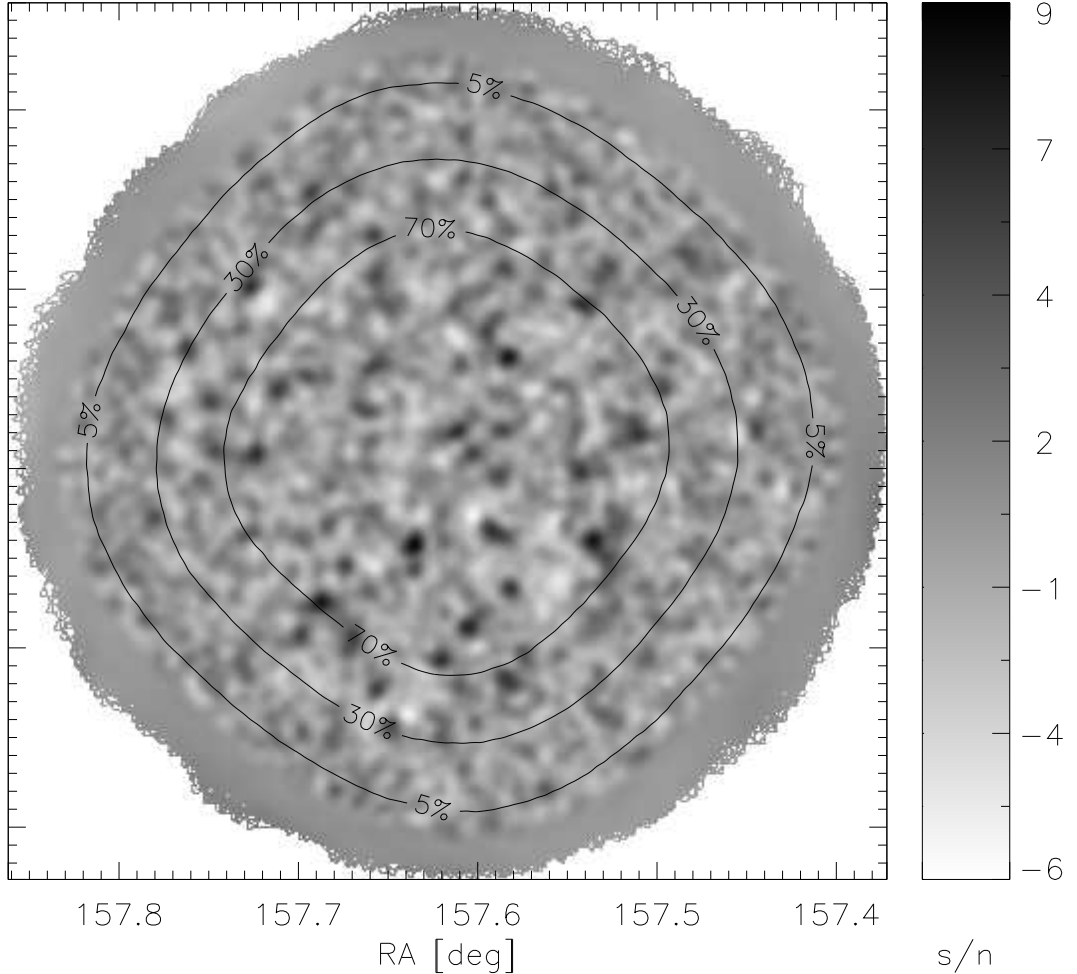


Fig. 5.— The filtered signal-to-noise map. The 5%, 30%, and 70% coverage contours are overlaid.

In terms of weighting a fit according to all possible sources of uncertainty, the filtering technique presented in section 4 (filter A) is clearly closer to optimal than the standard matched filter discussed in section 3 (filter B), especially in the presence of significant coverage gradients. However, filter A contains a larger number of steps and computations. In terms of implementation, we have shown that the intricacies of filter A have virtually no effect on the total computing time devoted to a field. However, converting these additional steps into functional computer code does require more time and effort. Therefore, it is fair to ask *how much* better filter A performs compared to filter B in a practical application. Of course, the answer depends on the specifics of the application. For instance, the acuteness of coverage gradients in a map will be important for this assessment. We

will try to answer this question in the context of the typical AzTEC map that we have used here for illustrative purposes.

First, when a fit is closer to optimal, it yields smaller error bars on the fit parameters than a less optimized fit. A fair way of comparing error bars here is to compare the T maps (generated using noise realizations, as described in section 5.3) that result from the two filters. We find that the post-filter coverage ($1/n_p^2$) is on average 5-6% better with filter A than with filter B over most of the map, including the $> 70\%$ coverage region. Thus, filter B is more prone to error, which in this context leads to higher rates of false detections as well as non-detections of truly significant sources.

To form an impression regarding these rates, we summarize in table 1 the results obtained with each filter in different regions of the map. The “source candidates” reported in table 1 are defined to be peaks in the signal-to-noise map (Fig. 5) that exceed a value of 3.5. This is a good choice of threshold because, on average, pure noise realizations yield only ~ 1 noise peak with signal/noise > 3.5 over the entire map. Given the non-uniformity of coverage within the 30-70% region, it is not surprising that filter B “detects” 3 peaks that are judged by filter A (more accurately) to have $s/n < 3.5$. It is interesting that even in the $> 70\%$ coverage region, the source lists generated with filters A and B differ by two sources. The results reported in table 1 provide some grounds for the reader to gauge whether the additional work involved in implementing filter A is worth the effort. Of course, the optimal filter presented here will be most useful when the coverage of a field has large non-uniformities, unlike in the example field used here.

Thus far, many AzTEC publications only provide source lists from map regions with $> 70\%$ coverage. As the optimal filter presented here is geared to handle coverage non-uniformities, we believe that the reliable region for source extraction can be extended to lower coverage thresholds. As mentioned in section 5.3, regions with $< 5\%$ coverage should be excluded. In fact, it may be best to exclude a larger region. For instance, not all elements of the detector array have imaged the outer regions of the map. Therefore, those regions may be systematically biased in some way compared to regions viewed and appraised by all detectors. Given that the AzTEC-ASTE detector

Table 1. The number of peaks with $s/n > 3.5$ found with the fully optimal filter of section 4 (filter A) and the standard matched filter of section 3 (filter B) are reported. The last column gives the number of “sources” found with filter B but not with filter A. The 30-70% coverage region has an area that is 75% of the $> 70\%$ coverage region.

Map region	Filter A	Filter B	B but not A
$> 70\%$ coverage	31	31	2
30-70% coverage	15	18	3

array has a foot print on the order of $480''$, it is reasonable to leave out a border of half that size, and use the $> 30\%$ coverage region of Fig. 5 for source identification. This represents an increase of 75% in the source extraction area. Table 1 shows that $\sim 50\%$ more source candidates can be obtained this way.

The optimal filter described here is now part of the AzTEC data analysis pipeline, which may be downloaded (by anonymous ftp) at <http://www.astro.umass.edu/aztec/Software/software.html>. Within this suite, the particular IDL routine that implements the optimal filter is *aztec_adapative_wiener_filter.pro*.

This work has been funded, in part, by NSF grant AST-0907952. KSS is supported by the National Radio Astronomy Observatory, which is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

A. Appendix: Fourier transform conventions

In this work, $\text{FT}[g]$ denotes the discrete 2-dimensional Fourier transform of the function $g(\mathbf{x})$. Following the convention used by IDL and other high-level programming languages, this Fourier transform is defined as

$$\tilde{g}(\mathbf{k}) = \tilde{g}(k_a \hat{i} + k_b \hat{j}) = \frac{1}{N_x N_y} \sum_{m=0}^{N_x-1} \sum_{n=0}^{N_y-1} g(x_m \hat{i} + y_n \hat{j}) e^{-2\pi j a m / N_x} e^{-2\pi j b n / N_y}, \quad (\text{A1})$$

where N_x and N_y are the number of pixels in the x and y dimensions. Therefore, $N_x N_y = N_{\text{pixel}}$. The x_m and y_n above, take on the x and y values of all the pixel centers. The a and b above assume integer values in the range $[0, N_x - 1]$ and $[0, N_y - 1]$ respectively. The discrete points on the reciprocal plane, where \tilde{g} is defined are specified by

$$k_a = \frac{1}{\Delta x} \left(\frac{1+a}{N_x} - \frac{1}{2} \right) \quad (\text{A2})$$

$$k_b = \frac{1}{\Delta y} \left(\frac{1+b}{N_y} - \frac{1}{2} \right), \quad (\text{A3})$$

where $\Delta x = \Delta y$ is the pixel size. The corresponding inverse Fourier transform, denoted $\text{IFT}[\tilde{g}]$, is defined as

$$g(\mathbf{x}) = g(x_m \hat{i} + y_n \hat{j}) = \sum_{a=0}^{N_x-1} \sum_{b=0}^{N_y-1} \tilde{g}(k_a \hat{i} + k_b \hat{j}) e^{2\pi j a m / N_x} e^{2\pi j b n / N_y}. \quad (\text{A4})$$

According to these conventions, the discrete convolution theorem takes on the form

$$\sum_l h(\mathbf{x}_l - \mathbf{x}_p) g(\mathbf{x}_l) = N_{\text{pixel}} \sum_a \tilde{h}^*(\mathbf{k}_a) \tilde{g}(\mathbf{k}_a) e^{2\pi j \mathbf{k}_a \cdot \mathbf{x}_p}, \quad (\text{A5})$$

when h and g are real functions. In Eq. A5, l is an index over all pixels of the map rather than a single dimension and a , similarly, is an index over all sampled points on the 2-d reciprocal plane.

REFERENCES

- Austermann, J. E. et al. 2010, MNRAS, 401, 160, 0907.1093
- Barnard, V. E., Vielva, P., Pierce-Price, D. P. I., Blain, A. W., Barreiro, R. B., Richer, J. S., & Qualtrough, C. 2004, MNRAS, 352, 961, arXiv:astro-ph/0405156
- Barreiro, R. B., Sanz, J. L., Herranz, D., & Martínez-González, E. 2003, MNRAS, 342, 119, arXiv:astro-ph/0302245
- Baugh, C. M., Lacey, C. G., Frenk, C. S., Granato, G. L., Silva, L., Bressan, A., Benson, A. J., & Cole, S. 2005, MNRAS, 356, 1191, arXiv:astro-ph/0406069
- Bertoldi, F., et al. 2007, ApJS, 172, 132
- Béthermin, M. et al. 2012, ApJ, 757, L23, 1208.6512
- Blain, A. W., Smail, I., Ivison, R. J., Kneib, J.-P., & Frayer, D. T. 2002, Phys. Rep., 369, 111, astro-ph/0202228
- Borys, C., Chapman, S., Halpern, M., & Scott, D. 2003, MNRAS, 344, 385, arXiv:astro-ph/0305444
- Chapin, E. L. et al. 2011, MNRAS, 411, 505, 1003.2647
- Coppin, K., et al. 2006, MNRAS, 372, 1621, astro-ph/0609039
- Downes, T. P., Welch, D., Scott, K. S., Austermann, J., Wilson, G. W., & Yun, M. S. 2012, MNRAS, 423, 529, 1103.3072
- Ezawa, H. et al. 2008, SPIE Conf. Series, 7012, 6E
- Humphrey, A. et al. 2011, MNRAS, 418, 74, 1107.3120
- Perera, T. A. et al. 2008, MNRAS, 391, 1227, 0806.3791
- Scott, K. S. et al. 2008, MNRAS, 385, 2225, 0801.2779
- Scott, K. S. et al. 2012, MNRAS, 423, 575, 1203.2609
- Stetson, P. B. 1987, PASP, 99, 191
- Tegmark, M., & de Oliveira-Costa, A. 1998, ApJ, 500, L83, arXiv:astro-ph/9802123
- Valiante, E., Lutz, D., Sturm, E., Genzel, R., & Chapin, E. L. 2009, ApJ, 701, 1814, 0906.4110

- Vieira, J. D. et al. 2010, *ApJ*, 719, 763, 0912.2338
- Vio, R., Andreani, P., & Wamsteker, W. 2004, *A&A*, 414, 17, arXiv:astro-ph/0305246
- Wei, A. et al. 2009, *ApJ*, 707, 1201, 0910.2821
- Wilson, G. W. et al. 2008, *MNRAS*, 386, 807, 0801.2783
- Younger, J. D. et al. 2007, *ApJ*, 671, 1531, 0708.1020
- Younger, J. D. et al. 2009, *ApJ*, 704, 803, 0909.3299
- Zeballos, M. et al. in prep., *MNRAS*